Nudges, Agency, Navigability, and Abstraction: A Reply to Critics

Preliminary draft 3/23/2015
For Review of Philosophy and Psychology, Special Issue on Nudges

Nudges, Agency, Navigability, and Abstraction: A Reply to Critics

Cass R. Sunstein*

Abstract

This essay, for a special issue of the Review of Philosophy and Psychology, responds to ten papers that explore the uses and limits of nudges and choice architecture. The essay has three general themes. The first involves the objection that nudging threatens human agency. My basic response is that human agency is fully retained (because nudges do not compromise freedom of choice) and that agency is always exercised in the context of some kind of choice architecture. The second theme involves the importance of having a sufficiently capacious sense of the category of nudges, and a full appreciation of the differences among them. Some nudges either enlist or combat behavioral biases but others do not, and even among those that do enlist or combat such biases, there are significant differences. The third general theme is the need to bring various concerns (including ethical ones) in close contact with particular examples. A legitimate point about default rules may not apply to warnings or reminders. An ethical objection to the use of social norms may not apply to information disclosure. Here as elsewhere, abstraction can be a trap. We continue to learn about the relevant ethical issues, about likely public reactions to nudging, and about differences across cultures and nations. Future progress will depend on a high level of concreteness, perhaps especially in dealing with the vexing problem of time-inconsistency.

Nudges are interventions that steer people in particular directions but that also allow them to go their own way (Thaler and Sunstein 2008; Sunstein 2015a). A reminder is a nudge; so is a warning. A GPS nudges; a default rule nudges. To qualify as a nudge, an intervention must not impose significant material incentives (Sunstein 2014). A subsidy is not a nudge; a tax is not a nudge; a fine or a jail sentence is not a nudge. If an intervention imposes significant material costs on choosers, it might of course be justified, but it is not a nudge. Some nudges work because they inform people; other nudges work because they make certain choices easier; still other nudges work because of the power of inertia and procrastination.

A nudge might be defended on the ground that it either enlists or helps counteract a behavioral bias of some kind (such as present bias, availability bias, or unrealistic optimism). For example, some information campaigns attempt to counteract present bias

* Robert Walmsley University Professor, Harvard University.

by encouraging people to focus on the long-term. But a behavioral bias is *not* a necessary condition for a nudge. Disclosure of information can be helpful even in the absence of any bias (because unbiased people may lack information), and while inertia might fortify the effect of a nudge, it would be highly inaccurate to say that the basic point of nudging is "to exploit human biases." A reminder may respond to a bias (inattention), but it does not exploit a bias. A default rule simplifies life and might therefore be desirable whether or not a behavioral bias is involved. A GPS is certainly a nudge; it tells you how you can best get to your preferred destination, but it does not impose any sanction or costs if you refuse to do what it says. A GPS is useful even for people who do not suffer from any kind of behavioral bias.

As the GPS example suggests, many nudges have the goal of *increasing navigability* – of making it easier for people to get to their own preferred destination. Such nudges stem from an understanding that life can be simple or hard to navigate, and a goal of helpful choice architecture is to promote simple navigation. A GPS does not undermine human agency; it promotes it. The same point holds many of the most impressive strategies for facilitating interactions between human beings and machines. (Apple's Steve Jobs was a superb choice architect.) To date, there has been far too little attention to the close relationship between increased navigability and (good) nudges. Insofar as the goal is to promote navigability, the ethical objections are greatly weakened and might well dissipate (at least if we trust those who claim that they are seeking to promote navigability).

It is essential to see that some form of choice architecture is inevitable. Human beings cannot wish it away, however much they are committed to freedom and autonomy. Any store has a design; some products are seen first, and others are not. Any menu places the options at various locations. Television stations come with different numbers, and strikingly, lower numbers are better, even when the costs of switching are vanishingly low; people are more likely to choose a station numbered 2 or 3 than one numbered 150 or 200. Any website has a design, which will affect what and whether people will choose. Streets, street signs, computers, cell phones, and ballots offer choice architecture of their own.

Nor can public officials avoid nudging. A nation that respects freedom of speech and freedom of religion, or that is committed to human dignity, will nudge people by virtue of that very fact. Constitutional law itself has an *expressive function*, which influences people even when it does not coerce. A Bill of Rights imposes prohibitions on government, to be sure, but it also operates at least a bit like a GPS, providing public officials and citizens with important guidance about defining social commitments. Insofar as a Constitution safeguards freedom of speech, private property, or human dignity, it will help create a kind of choice architecture, and it will nudge.

Any government, even one that is or purports to be firmly committed to active choosing, free markets, and the idea of laissez-faire, will almost inevitably provide a set of prohibitions and permissions, including a number of default entitlements, establishing who has what before bargaining begins. The rules of contract, property, and tort provide a

form of choice architecture for social ordering; they contain default rules as well as bans as authorizations. (Some prominent critics of nudging and choice architecture – especially skeptical economists and psychologists -- neglect this point, which is familiar to lawyers.) Such rules facilitate, signal, and prohibit, among other things – and this is quite important – through *extinguishing some of people's self-help remedies*, by forbidding people from "taking the law into their own hands." If someone has violated your contractual rights, or intruded on your property, or asserted property rights against you, you must resort to legally approved channels. This requirement is a mandate, and no mere nudge, but it too has an expressive function, helping to create preferences and values.

Default rules are omnipresent, and they help constitute the law of contract, much of which has the following form: *Unless the contracting parties say otherwise, the rules that govern their relationship will be as follows; but the parties are permitted to say otherwise*. People's legal relationships with their employer, their mortgage provider, their rental car company, their credit card company, and even their spouse and their children consist in large part of default rules. People can alter the default rules, to be sure, but often they do not, not least because of the power of inertia and suggestion (Sunstein, 2015b). The default rules will often operate like a GPS, or even help to shape preferences and values.

To some people, this is a disturbing or even threatening fact: They believe that people should be able to organize their lives as they like, and they are not at all enthusiastic about the idea that significant aspects of their lives are organized by default rules, which they did not themselves select, and which might well come from the practices, judgments, or wishes of other people. (As we shall see, many people object to nudges because they prize individual agency.) Nonetheless, that organization is in place. Moreover, it is true that some default rules are a product of traditions, customs, spontaneous orders, and invisible hands – a comfort to some people who especially distrust public officials. But it would be extravagant to say that all (or most) of them are. And even if they are, they will nudge individuals who live with them, and it takes real work to transform them into law, where they will have significant effects.

Whenever a government has offices, designs highways, or maintains a court system and official websites, it will nudge. A government that maintains an educational system, or that imposes various incentives and requirements on educators, will certainly nudge. Any educational system is replete with nudges. It is true that choice architecture can (and should) insist on a great deal of room for freedom of choice. It should go without saying that active choosing and learning are important. It is also true that choice architects can at least aspire to important kinds of neutrality, fit for a free and self-governing society -- including, for example, neutrality among religions, between men and women, and across political views. But choice architecture itself is inevitable. It cannot be wished away.

I am most grateful to the contributors to this issue for their many valuable and illuminating thoughts on the subject of nudges and choice architecture. In this response, I

offer some remarks on their various arguments and suggestions. Insofar as I have general themes, they are threefold. The first and most important involves the widespread concern that nudging threatens human agency. My basic response is that when nudges are in place, human agency is retained (because freedom of choice is not compromised) and that agency always takes place in the context of some kind of choice architecture. The second general theme involves the importance of having a sufficiently capacious sense of the category of nudges, and an appreciation of the differences among them. As noted, some nudges combat behavioral biases but others do not, and even among those that do combat such biases, there are significant differences. The third general theme is the need to bring various concerns (including ethical ones) in close contact with particular examples. A legitimate point about default rules may not apply to warnings or reminders. An ethical objection to the use of social norms may not apply to information disclosure.

There is a close connection among the three general themes. If our concern is human agency (on grounds of either autonomy or welfare), and if we seek to avoid interventions that threaten it, our principal target should be mandates and bans. Nonetheless, it is true that the idea of nudging – and indeed the very word – might be taken to suggest impositions by some kind of elite, determined to steer people in the directions that it prefers. We can readily imagine nudges that would run afoul of this objection. But in a nation that is committed to both individual liberty and social welfare, those nudges are unacceptable. Desirable nudges should undermine neither autonomy nor welfare. As we shall see, they can promote both values; indeed, they might be indispensable for them.

Here as elsewhere, abstraction can be a trap. People tend to get carried away with them. Abstract concerns are important, but there is a real risk, which is that they will create confusion unless they are brought to bear on particular practices. In his Marginalia on Sir Joshua Reynolds, William Blake wrote, "To Generalize is to be an Idiot To Particularize is the Alone Distinction of Merit – General Knowledges are those Knowledges that Idiots possess." Characteristically, the great poet put it much too strongly. But he had a point.

Hagman, Andersson, Vastfjall, and Tinghog

What do people actually think about nudges? Do they object on ethical grounds? Do they distinguish among nudges? On what grounds? Of course answers to such questions would not resolve the ethical questions. People might reject approaches that they ought on reflection to endorse, or vice-versa. But information about people's reactions is independently interesting, not least because it says something about likely democratic responses, and also because it might inform ethical thinking as well (Felsen et al. 2013).

Surveying 952 people in Sweden and the United States, Hagman, Anderson, Vastfjall and Tinghog produce a striking and important conclusion: In general, strong majorities of both Swedes and Americans are positively inclined toward nudges. Consider, for example, these:

1. Avoiding tax evasion. *Many countries have a problem with its citizens not paying taxes, which costs society a considerable amount of money. Some countries have therefore started to send out information to the taxpayers with the encouraging message "To pay your taxes is the right thing to do". The idea with this intervention is to give tax evaders a bad conscience and therefore increase their motivation to pay their taxes.*

2. Smoking discouragement. *Smoking often leads to addiction and has a negative effect on the health of the individual. To more clearly show the negative effects of smoking, many countries have started to add deterrent pictures on the cigarette packages. These images display damaged organs that can be a consequence of long term smoking. The idea with this intervention is to discourage people to start smoking and motivate people that are smokers to quit.*

3. Cafeteria. *Overconsumption of calorie rich food can lead to a deteriorating health. In an attempt to get their employees to eat healthier, a company rearranged its cafeteria. Healthy food was placed at eye-level and easily available for the visitors of the cafeteria. Unhealthy food, such as candy or snacks was placed behind the counter to make them less visible and accessible for the visitors in the cafeteria. The idea with this intervention is to encourage the consumption of healthier alternatives to improve the health of the employees.*

Hagman et al. find that over 80 percent of both Swedes and Americans find the tax evasion policy acceptable, and they find comparable numbers for both smoking discouragement policy and cafeteria re design (81 percent of Swedes, 72.6 percent of Americans) and the third (82.6 percent of Swedes, and 76.4 percent of Americans as well. Consistent with expectations, Swedes are somewhat more enthusiastic than Americans about nudges, but only two of the eight nudges fail to attract majority support in either country, with 42.9 percent and 45.7 percent of Americans (but over 60 percent of Swedes) favoring these:

4. Organ donation. *There is currently a lack of organ donors in many countries. In some places, to become an organ donor the individual has to make an active choice and register as an organ donor with the appropriate authority. If no choice is registered, the individual is assumed to be unwilling to donate in event of an accident (so called Opt-In). In previous surveys most people report that they are willing to be an organ donor but have not registered.*

One way to increase the number of organ donors could be to automatically enroll people as organ donors unless otherwise specified (so called Opt-Out). In other words, it is up to the individual to register at the appropriate authority if they are unwilling to donate their organs in the event of an accident. The aim with this intervention (Opt-Out) is to increase the number of organ donors.

5. Climate compensation. *Carbon dioxide emissions in connection with flying have a negative effect on the environment. To compensate for this, there is usually a voluntary fee that travelers can add to the final price. The money from this fee goes to projects to reduce emissions of carbon dioxide to a corresponding level of the emission caused by the flight. To increase the number of travelers that choose to pay the climate compensation fee, it can automatically be added to the final price. Then, if a traveler does not want to pay the fee, the traveler instead has to make an active choice not to pay the fee (also known as Opt-Out). The idea with this intervention (Opt-Out) is to increase the number of travelers that compensate for climate.*

In the United States, majority rejection of these forms of choice architecture – and significant opposition in Sweden as well – likely stems from this judgment: *choice architects should not use people's inertia or inattention against them*. For decisions that have a significant degree of moral sensitivity (organ donation) or cost (climate change compensation), many people reject a default and would undoubtedly favor active choosing. The apparent idea – for which more empirical testing would be desirable -- is that if a default rule would lead people to end up with an outcome that is morally troubling (to them) or expensive (for them), that rule is objectionable and active choosing is much better.

That lesson is a significant one, but the more important finding is the apparently widespread endorsement of nudges, whether the goal is to protect third parties (as in the case of tax evasion) or the self (as in the case of smoking discouragement). Not surprisingly, Hagman et al. also find that those with an individualistic worldview are (somewhat) less likely to embrace nudges. More strikingly, they find that respondents with a strong preference for analytical thinking are less likely to see nudges as intruding on freedom of choice.

For those who are focused on ethical issues, Hagman et al. provide a valuable reality check, offering significant insights into people's reactions to a range of nudges. The Swedish-U.S. differences are interesting even in the midst of the general agreement between people in the two nations. It would of course be valuable to test a wider array of nudges and to see what kinds of division emerge. We could explore whether there are systematic differences between "harm-to-self" nudges and "harm-to-others" nudges. We could also see whether people reject particular categories of nudges -- for example, those that seem to involve particularly personal choices. My hunch is that if we tested a large category of nudges, some interesting ethical distinctions would emerge -- but these cannot be easily anticipated in advance. On the underlying issues, Hagman et al. have made an important start (see also Felsen et al. 2013).

Mills

What is the relationship between autonomy and nudging? Some people believe that it is uneasy, and that nudges can compromise autonomy, rightly understood. In his

highly illuminating essay, Chris Mills argues that that this belief is far too crude. Many nudges can promote autonomy, which Mills understands to be "the capacity for an individual to determine and pursue her own conception of the good according to her own will." Mills demonstrates that if the purpose and effect of a nudge are to facilitate an individual's pursuit of her own goals, the autonomy objection is much weakened – and indeed, that certain kinds of nudges can serve to promote autonomy. He shows that techniques are both needed and available to ensure that we have access to people's authentic preferences and values. If the cost of opt-out is low, and if publicity and transparency are guaranteed, then there is far less threat to autonomy. Mills contends that personalized default rules, active choosing ("choice prompts"), and framed provision of information can meet these requirements. In this way, Mills rightly defuses the autonomy objection by explicitly connecting certain nudges with the idea of agency, which they promote rather than undermine.

Mills is also concerned about "epistemic paternalism," which occurs when choice architects fail to respect people as competent choosers, and which might seem to be an especially serious problem when they harness, appeal to, or enlist behavioral biases. Selective framing of information can be "an epistemic threat to the autonomy of the subject." Alert to the trap of abstraction, Mills contends that to know whether any apparent epistemic threat is real, we need to investigate the particular form of choice architecture. Active choosing, for example, need not be a form of paternalism at all, and a well-designed default rule ought not to count as an objectionable form to the extent that it does not increase the costs of inquiry and allows the subject freely to opt out. It is relevant here that often the only realistic choice is between an opt-in or an opt-out default; it does not make a great deal of sense to object that both forms violate autonomy. (Recall, however, that active choosing is sometimes feasible and best.)

Mills' careful analysis seems to me broadly convincing, and it stands as a valuable step forward. A virtue of his analysis is its clarity about when nudges might threaten autonomy. If choice architects do not respect subjects' ends, or if opt-out is seriously limited, autonomy is genuinely at risk. Mill might have added that when people do not want to choose, it can be paternalistic to require them to do so (Sunstein 2015b). A remaining question is how the analysis applies to a much larger set of nudges. Consider, for example, graphic health warnings, designed to reduce smoking; cafeteria design, meant to reduce unhealthy eating; invocation of social norms, meant to promote energy conservation; and text reminders, meant to encourage people to pay their bills. In my view, all of these can respect autonomy insofar as they preserve freedom of choice and do not compromise agency – but more work, both conceptual and empirical, would be valuable on that large question (for relevant discussion, see Wilkinson 2013; Sunstein 2015c). There is also a question about how to assess nudges that undermine autonomy but promise to increase welfare; answering that question would require us to take a stand on some foundational questions.

Whitman and Rizzo

A central goal of nudging is to "make choosers better off, as judged by themselves" (Thaler and Sunstein 2008: 5). (I am putting to one side the case of third-party effects, taken up below.) Notwithstanding that goal, Whitman and Rizzo do not believe that it is possible to establish normative standards for welfare-enhancing individual behavior. That is an extreme claim, and I wonder whether Whitman and Rizzo actually believe it. For orientation, consider the following cases:

1. Jones is asked to make a choice between two identical radios. One costs more. He chooses the more expensive one.
2. Jones orders fish tacos at a restaurant. He used to like fish tacos, but he hasn't recently. (He forgot that fact; he ordered them out of habit.) He doesn't enjoy the fish tacos.
3. Jones often gets lost driving to local restaurants. He has a poor sense of direction. His girlfriend gets him a GPS, and he no longer gets lost. He's glad he has the GPS.
4. Jones is asked to make a choice between two health insurance plans. One dominates the other: It is better along several dimensions and worse along no dimension. Jones chooses the dominated plan.
5. Jones is buying a car. The first option costs very slightly less than the second but has terrible fuel efficiency, so much so that after six months, the second option would save him money. (He expects to own the car for at least five years.) He likes the two cars the same. He chooses the first option, because he pays no attention to the fuel economy of the cars.
6. Because he is busy and inattentive, Jones, who is poor, is often late paying his credit card bills. If he received text messages from his credit card company, he would make his payments on time. His credit card company does not send him text messages.
7. Jones fails to sign up for a retirement plan. He is aware that his employer has such a plan, and he thinks that enrolling would be an excellent idea, but he keeps thinking that he will sign up "next month." Next month turns out to take a while. He does not starting saving for five years.

Whitman and Rizzo do not focus on problems of this kind. They contend that for libertarian paternalists, decision-making failures are demonstrated by departures from the neoclassical model of rationality. In their view, the behavioral paternalist case turns out to depend on "the *normative* strength of the neoclassical rationality axioms that are violated by decision-making anomalies" (emphasis in original). As they show, however, those axioms were originally designed as positive statements lacking normative content, and in their view, they have questionable normative status. Whitman and Rizzo think that an agent who lacks completeness and intransitivity need not be described as "irrational." Among other things, people might form their preferences during the process of choice, which means that analysts lack an Archimedean point – preexisting preferences – from which to judge outcomes. (Behavioral economists agree, see Goldin, 2015.)

Whitman and Rizzo also think that if we find inconsistent choices, we will not know which to favor on welfare grounds. A person might have one rate of time preference in January and another in June, showing both high and low levels of patience; which should we honor? A person may behave in one way in a hot state (hungry) and in another way in a cold state (satiated); which behavior should be taken as normative? With a default rule in favor of vacation time, people demand more to give up vacation time than they would pay to obtain it if they did not have it by default. How do we know, from this evidence, which default rule is best?

Whitman and Rizzo raise good questions, on which far more work needs to be done. It is surely right to question whether we should always prioritize the preferences of people in a cold state. With respect to default rules, recent work is trying to provide some answers (see the excellent discussion in Goldin 2015), not by emphasizing inconsistency itself, but by developing principles for assessing the welfare issue. (Those principles do not invoke the neoclassical axioms.) It seems best to take Whitman and Rizzo as laying down a challenge for more work in this vein, rather than as a basis for skepticism about the very possibility of developing normative standards. Whitman and Rizzo are entirely right to emphasize the importance of epistemic humility on the part of choice architects, but if people suffer from serious self-control problems – if they are making themselves sick or dead, are aware of that fact, and wish it were otherwise – perhaps we should not much trouble ourselves about the supposed indeterminacy of welfarist criteria.

Whitman and Rizzo assume far too close a link between the neoclassical axioms, and in particular the problem of inconsistency, and the welfarist arguments for nudges. Recall the seven cases with which I began. To be sure, we can complicate those cases in such a way as to make the welfare assessment more difficult. But some cases are easy, and difficult cases present an opportunity for better thinking, not for skepticism about the whole project.

Guala and Mittone

Some nudges are designed to help choosers themselves; they might be paternalistic. Other nudges are designed to respond to market failures, as, for example, by reducing externalities. Environmental nudges fall in this category (Bubb and Pildes 2014; Sunstein and Reisch 2014). In the latter case, there may be a convincing welfarist argument for nudges, and cost-benefit analysis is highly relevant. Indeed, mandates, and not merely nudges, might turn out to be justified.

Sounding a lot like Whitman and Rizzo, Guala and Mittone contend that in important cases, welfarist arguments on behalf of nudges do not work, at least not if we are focused only on the welfare of choosers. The reason is that we do not know how to make the relevant assessment. Suppose that young John does not want to save for retirement, likes to eat a lot of high-calorie food, or despises exercise; suppose that as a result, old John is poor, fat, and unhealthy. Suppose too that young John is happy if he does not save, eats a lot, and does not exercise, but that old John is miserable as a result of young John's choices. Guala and Mittone think that on welfare grounds, we cannot

decide which John to favor. In the face of this difficulty, they argue that we should instead embrace a "political" approach that uses nudges "to protect *other* people from the damage that might be caused by irresponsible individuals" (emphasis in original). They insist that we should not rely on "welfarism," which is in their view "a red herring," but should instead invoke "a political justification."

I am not sure exactly what this means. Begin with a semantic clarification: By a political justification, Guala and Mittone mean, in part, to invoke a standard welfarist argument, based on the presence of externalities. They mean to suggest that when a chooser inflicts costs on third parties (including taxpayers), it is legitimate for the state to intervene. If John's obesity would inflict such costs, then on standard welfarist grounds, John cannot object to an intervention that is designed not protect John, but those whom he is injuring.

The first point to make about this argument is that it is contingent on an empirical claim, which is that those who run health risks do, in fact, end up imposing (net) costs on third parties. Suppose that obese people or smokers die young, and for that reason *reduce* overall costs on third parties. If so, the externalities argument loses its empirical foundation. But Guala and Mittone offer another point, which is that citizens might believe that even if the costs of myopia are not externalized, they have a moral duty to help those who have made mistakes in the past. Citizens might think that premature deaths are bad, or wrong, and that nudges, reducing or eliminating those deaths, are justified because they reduce or eliminate the probability that citizens will feel (morally) obliged to intervene.

Taken in a certain way, however, this argument proves too much. In their account, citizens are saying that they can nudge Sarah in certain directions at Time 1 because if they do not, they will feel a moral obligation to intervene in Sarah's life at Time 2 (and thus face costs at that time). Suppose that Sarah wants to be a poet (which means that she might well end up poor), or that she is attracted to women (which means – let us suppose - that citizens will disapprove, on moral grounds, of her romantic choices), or that she wants to be an astronaut (which means that she will run some serious risks), or that she embraces a highly unpopular religion (which means – let us suppose – that citizens will want to intervene, soon or eventually, on moral grounds). It cannot be the case that a nudge is justified merely because it reduces the likelihood that a majority of citizens will feel morally obliged to intervene.

At a minimum, they must, in fact, be morally obliged to intervene, and not merely believe that they are, and with that qualification, all of the moral questions are put right back on the table. Recall that the issue is this: May citizens nudge Sarah at Time 1 because they will feel obliged to intervene at Time 2? Something like a positive answer might be welfarist: They may nudge Sarah at Time 1 if and because Sarah will have a better life as a result. Another kind of positive answer would invoke autonomy: They may nudge Sarah at Time 1 if and because Sarah will have more autonomy as a result. (Recall Mills.) But in either case, it is not relevant that citizens might feel morally

obliged to intervene at Time 2. I do not think that Guala and Mittone offer a convincing political argument for nudges.

In my view, Guala and Mittone reject the welfarist argument for nudges far too quickly. In some cases, people make mistakes because they lack information or because they cannot process it properly (Bhargava et al. 2015), or because they are confused, inattentive, or forgetful (Cadena and Schoar 2011; York and Loeb 2014). In such cases, there is no problem in concluding that a nudge can be helpful. We have seen that the analysis may be more difficult when people face self-control problems or when choices increase welfare at Time 1 but reduce it at Time 2 (and I will return to this problem). But even in such cases, a welfarist assessment might turn out to be possible. We know enough to know that if they are suitably designed, automatic enrollment programs, in the context of savings, can do a great deal of good (Thaler and Bernartzi 2013). We know that with some kinds of choice architecture, people's choices reflect their true preferences better than with other kinds of choice architecture; the best kinds might well include a nudge (Alcott and Sunstein 2015). Once more: In the face of time inconsistency, there is a pressing need for more conceptual and empirical work, but there is no reason to think that we will fail to make progress, even on the hardest questions.

Gigerenzer

There is no opposition between education and nudges, any more than there is an opposition between education and occupational safety regulation (which contains numerous nudges), or between education and contract law (which also contains numerous nudges), or between education and the criminal law. Those who favor education and those who favor nudges have no quarrel. Moreover, an insistence on the general usefulness of heuristics, and an optimistic view of human rationality, are entirely compatible with a receptive attitude toward nudges and good choice architecture. (Recall the trap of abstraction.) Indeed, those who emphasize the value of heuristics are often at pains to identify, and to call for, forms of choice architecture that do not trick or confound people's intuitions.

Gerd Gigerenzer is right to say that heuristics often work well; that is why they exist. Nor does it make much sense to deny the potential value of education. After all, disclosure of information is a nudge, and it is certainly educative. Education as such might itself be counted as a nudge insofar as it is designed to maintain freedom of choice while steering people in certain ways (if only to exercise their own agency). To be sure, there are serious empirical questions about the value and efficacy of education in some settings that have concerned behavioral scientists (on financial education, see Willis 2013). But in important contexts, Gigerenzer's optimism might prove justified.

We should not say "education yes, nudges no"; both are indispensable. If people can be made "risk savvy," all the better. It would be most surprising if Gigerenzer were, on reflection, opposed to default rules as such. Life and law are not possible without them, so we might as well have good ones. Importantly, there are serious discussions to be had about the choice between active choosing (perhaps accompanied by education)

and default rules (Sunstein, 2015b), and sometimes the best choice architecture calls for the former. If Gigerenzer is concerned about the power of government (and everyone should be), his primary objection should not be to choice-preserving nudges, but to mandates and bans. But it would not make a lot of sense to say that because education can work, we should eliminate energy efficiency rules, food safety rules, and criminal punishment.

Gigerenzer is wrong to suggest that those who believe in nudges have a "dismal picture of human nature." It is possible to have a sunny view of human nature and to believe that good choice architecture and nudges can help. It is also possible to insist that human beings are human rather than divine, and subject to predictable biases, without having a "dismal picture of human nature." We can acknowledge that some people procrastinate, that others are unrealistically optimistic, that others run serious health risks, and that others are discourteous or even violent, while also celebrating human rationality and decency (as manifested, in part, by laws, punishments, and systems of government, not excluding nudges).

Gigerenzer reports that those who favor good choice architecture indulge an "assumption that choice architects are benevolent." Nothing could be further from the truth (Jolls et al., 1998; Camerer et al., 2003; Thaler and Sunstein, 2008). A central reason for nudges, as opposed to mandates, is that choice architects are not always benevolent (Thaler and Sunstein, 2008; Sunstein, 2015b). Choice architecture is not avoidable, so it is pointless to try to wish it away. The risk of malevolent or ignorant choice architects argues in favor of default rules rather than mandates (Sunstein 2015b) and also for active choosing, which (again) is a form of choice architecture (ibid.). Potentially mistaken (or self-interested or venal) choice architects have long been, and continue to be, among the central concern of those who focus on the topic of choice architecture (Jolls et al., 1998; Sunstein, 2015b).

Gigerenzer has mounted innumerable attacks on the work of Nobel Prize winner Daniel Kahneman (and his late coauthor Amos Tversky, who would undoubtedly have shared the Nobel if he had lived). He repeats some of his longstanding claims here. Whatever their merits, those claims are largely irrelevant to the current topic. Even if we have an upbeat attitude toward heuristics, and conclude that Gigerenzer is correct on some or all of the psychological issues, we will hardly cease to be favorably disposed toward sensible default rules and good choice architecture. You can think that Gigerenzer is right on framing effects and rationality while supporting disclosure of information, simplification, warnings, reminders, and good default rules (on the role of such support in the Obama Administration, see Sunstein 2013).

It is true that Thaler and I have endorsed some psychological claims that Gigerenzer rejects (Thaler and Sunstein 2008), but most of the time, those academic debates need not be resolved in order to embrace nudges as policy instruments. A GPS is a good idea even if most people are excellent navigators. A reminder can be helpful even if people have a good understanding of risk. Here is a question that might reduce promote progress (while avoiding the trap of abstraction): If we think that heuristics work

exceedingly well, which concrete nudges, actually adopted on behavioral grounds by policymakers in (say) the United States or the United Kingdom, would we therefore reject? What would show up on any such list? I suspect that the list would not be long.

Nagatsu

Some nudges (like some social norms, Ullmann-Margalit 1977) serve to increase the voluntary provision of public goods. In the context of environmental harms, for example, a goal of good choice architecture is to solve a collective action problem; nudges might contribute to that goal. Payment of taxes is not exactly voluntary, but a well-functioning tax system depends on payments that people make without actual or threatened enforcement action, and nudges might be able to reduce the level of delinquency.

Michiru Nagatsu is concerned that some nudges might run into a objection from autonomy, because they "induce behavioral changes to which one's reasoning process is not responsive" (on this topic, see Sunstein 2015c). Nagatsu contends that in the case of social dilemmas, this objection does not have much force. Some people are conditional cooperators: They will cooperate if they think that a sufficient number of other people will do the same. Moreover, some nudges help to shift people from an "I-frame" (what should *I* do?) to a "we-frame" (what should *we* do?), thus helping to resolve a social dilemma.

Exploring Texas' successful anti-littering campaign ("Don't Mess With Texas"), Nagatsu suggests that it might have worked because it generated an expectation, among many people in Texas, that other people would refrain from littering. If so, the campaign did not damage to "autonomous agency in the sense of the capacity to reason." Alternatively, the campaign might have been a form of priming, focused on "Texas pride" and thus nudging Texans to adopt a we-frame. If so, autonomy might also have been preserved.

This is an impressively detailed and careful analysis, and it is convincing in its own terms, but we might question how often Nagatsu's argument, or some variation on it, is necessary to justify nudges. As we have seen (and I am sure that Nagatsu would agree), some nudges do not raise problems in terms of autonomy; consider information disclosure, warnings, and reminders. Mills' discussion helpfully shows, I think, that for many nudges, Nagatsu's argument is not required.

Nagatsu is also concerned about what he calls an objection from coherence, which arises when nudges actually produce inconsistent preferences. His point appears to be that even if good nudges steer a majority of agents towards *more* consistency (which is the goal), they can also *cause* inconsistency for some people. For example, nudges towards healthy food may produce inconsistency in people who would, all things considered, "genuinely prefer the current unhealthy food choice plus the risk of diseases and earlier death, to the current healthy food choice plus the prospect of living healthier and longer." We would have to specify an actual case to be sure, but in the abstract, the

concern cannot be ruled out of bounds. But it is at least a partial safeguard if freedom of choice is fully maintained, and if those who prefer the unhealthy choice can go their preferred route if they wish to do so.

Felsen and Reiner

Neuroscience is telling us a great deal about some of the sources of present bias, unrealistic optimism, probability neglect, and loss aversion (see, for example, Sharot 2010). In light of the growing research, Gideon Felsen and Peter Reiner contend that neuroscience can enable "us to make testable predictions about the effectiveness of nudges." For example, they urge that a nudge will be more effective if it influences decisions made with less in the way of conscious, top-down control. It is also possible that a nudge will have a larger impact if it affects people's sensory perceptions; reducing the aroma of brownies might have a bigger effect than moving them to a more remote location.

Felsen and Reiner also contend that neuroscience can illuminate normative questions. They emphasize the importance of distinguishing between higher-order desires (those involving fundamental goals) and lower-order ones (those involving physiological needs). Some nudges might promote a choice that is connected with higher-order desires and in that sense increase people's autonomy. (The point is correct and worth underlining.) They also urge that "external influences are the norm rather than the exception" and hence that "neuroscience suggests the degree to which our everyday decisions are autonomous – according to the consensus conception employed here – is limited."

Felsen and Reiner are cautious, and properly so, about offering broad conclusions about the relationship among neuroscientific findings, nudging, and choice architecture. If we understand behavior, we might know all that we need; the neural mechanisms might not much matter. Loss aversion is important whether or not it has identifiable neural foundations. But Felsen and Reiner are right to suggest that neuroscience might illuminate some of the ethical issues. When people are affected by learning about what other people do, is it because of an informational signal, producing higher-order processing, or is it because of some kind of immediate, automatic affective reaction? On one account, manipulation is distinctly associated with efforts to bypass people's reflective or deliberative capacities (for a review and a mixed verdict, see Barnhill, 2014; see also Sunstein, 2015c); neuroscience might be able to inform judgments about whether and when such bypassing is occurring.

I do not think that Felsen and Reiner would insist that once we identify neural mechanisms, we will necessarily have a clear sense of which nudges are most effective. A nudge that affects people's sensory perceptions (say, through aroma) might have only modest consequences as compared to an educative nudge (say, through disclosure of health-related information). The mechanism of an effect need not determine the magnitude of an effect. Nonetheless, it is surely true that some nudges that influence

lower forms of processing may be highly influential, again raising the question of manipulation (Barnhill, 2014).

Lepenies and Malecka

Some nations, including the United States, the United Kingdom, and Germany, have created behavioral insights teams of one or another kind, with a particular interest in empirical testing and choice architecture (Halpern, 2015). But to incorporate behavioral findings, no dedicated team is necessary. Some nations, including the United States and the United Kingdom, have had high-level officials with an interest in behaviorally informed approaches, with consequences for important legislation and regulation (see the discussion of automatic enrollment in Obama 2008 and the general overview in Sunstein 2013). Institutional design greatly matters; choice architecture is necessary for choice architects, in part to constrain them (Sunstein 2013).

Robert Lepenies and Magdalena Malecka distinguish between what they call an "individualistic approach," focusing on a nudger and a nudgee, and an institutional approach, focusing on "the legal and political institutions in which a nudge is embedded." They are right to notice that many nudges are not part of a legal system. A credit card company might send text reminders that a bill is due; a private university might automatically enroll students and staff in certain programs; a mortgage company might provide a simplified application. As a law professor, I would prefer not to say, as Lepenies and Malecka do, that nudges and the legal system are "estranged"; their relationship is far more congenial than that, and sometimes it is quite close. (For just one example, recall the presence of default rules in the law of contract.) But it is certainly true that much nudging occurs without the involvement of law, and in free societies, properly so.

Lepenies and Malecka also distinguish between "law-as-normative" and "law-as-instrumental." With the former, people see a requirement, prescribed by a legal norm, as a reason for action; if so, it must be "cognitively accessible." (An obvious example is a speed limit law.) With the latter, the goal is to change the context in which people make decisions, and law need not be cognitively accessible to be effective. (A default rule might be an example.) As Lepenies and Malecka put it, "the non-cognitive reaction of an agent is a sufficient condition for effective responding to the law." Lepenies and Malecka think that nudges fall in the category of law-as-instrumental, because they "are non-normative and they impact people's behavior in a non-cognitive way."

Lepenies and Malecka draw some disturbing conclusions from this categorization. Nudges reflect a view of human agency by which "individuals never deliberate about norms" and are restricted to emotions, sentiments, and automatic reactions. In their account, one result of nudging is to abandon the idea of influencing people with legal norms, and another is to deprive society of the ability to engage in self-legislation.

They fear that there "are very few safeguards to nudges today." In response, they embrace a panoply of new reforms. They argue for the creation a formal legal registry of

nudges, including (indeed "especially") those that are not part of the legal system. They think that because nudges suffer from "invisibility," they should have expiration dates. They want a "nudging ombudsman." More modestly, they think that graphic health warnings should include information about the legal source of such warnings. They argue as well that whenever citizens are nudged through default rules, they should be required to make an active choice between the default and opt-out.

These proposals seem to reflect a judgment that the globe is facing some kind of crisis of nudging, for which truly radical measures are required. To put it mildly, that seems a bit overstated. If the government is requiring disclosure of information, warnings, or particular default rules, safeguards should certainly be in place, including transparency and very possibly, an opportunity for public comment as well. Insofar as they come from government, nudges need not (and should not) suffer from any problem of invisibility; in the United States, for example, fuel economy labels, cigarettes warnings, energy efficiency labels, graphic health warnings, and automatic enrollment in retirement, school meals, and health insurance plans have been highly visible.

Everything that government does should be scrutinized, but if we want to impose an extra level of scrutiny, criminal law, which imposes real penalties and eliminates freedom of choice, would seem to be a better candidate than nudges. To be sure, government should identify the legal source of its actions (including use of graphic warnings), but expiration dates are often more trouble than they are worth. Just as we would not want expiration dates on laws that forbid murder, rape, and assault, or on the requirements of contract and property law, so we should not impose expiration dates on laws that call for disclosure or warnings. Default rules can be an excellent idea, and it would be far too cumbersome to force people, always, to make an active choice between a default and opt-out (Sunstein 2015b). In the context of retirement planning, for example, a system of simple defaults can work well. And it would be a lot more trouble to work with a computer or a cell phone if you were forced to decide, repeatedly and actively, between active choosing and reliance on some default.

In a free society, there is no need for a legal registry of nudges. For one thing, there are a lot of nudges out there (have a look at the television, or listen to the radio, or look at four random websites). For another, the very idea of a legal registry of nudges is not feasible. Any such registry would quickly become intolerably long – and from the standpoint of freedom, it would hardly be an unmixed blessing, because it would involve a high degree of legal intrusiveness into the private domain.

I am not sure what problem Lepenies and Malecka are trying to solve. The defining feature of nudges is not that they are invisible, but that they preserve freedom of choice. Many nudges are "cognitively salient" in their sense. It is true that default rules may work because people do not pay attention to them, but such rules cannot be avoided, and so we should not wish them away. Contract law could not exist without them. What Lepenies and Malecka call "the normative function of law" is important, but it would be extravagant to understand it in a way that would throw default rules into some kind of purgatory.

Kapsner and Sandfuchs

Do nudges interfere with privacy? Some of them certainly could. Imagine a default rule that says that unless you specifically indicate otherwise, all of your movements – online, in your city, as you travel – will be placed in the public domain. We could easily imagine forms of choice architecture that do not give privacy sufficient weight.

Kapsner and Sandfuchs are right to emphasize the importance of privacy and its connection with autonomy. They argue convincingly that to know how to nudge, choice architects might have to assemble a great deal of information, some of which people might want to keep private. If public officials are designing default rules that could compromise privacy, they should want to know whether the affected people (as individuals) care about privacy or not. For those who do care, such officials might adopt privacy-protective defaults; for those who do not, they might not. So far, perhaps, so good. But Kapsner and Sandfuchs make a clever (and to my knowledge original) objection, which is that *people might not even want government to know whether they care about privacy*.

In this light, Kapsner and Sandfuchs take objection to a thought experiment (of mine) in which a choice architect, whether public or private, has direct access to people's concerns and provides them with information about what they want. The thought experiment was never designed to suggest that such access would be a good idea, but Kapsner and Sandfuchs insist (quite rightly) that people might not want others – and especially not their government – to have access to those concerns. With respect to privacy, they worry in particular that if consumers receive messages about their electricity use, there might be a threat to privacy (other people might see the message), and that with automatic enrollment in an organ donation program, those who opt out are recorded as having done so.

Kapsner and Sandfuchs are correct to point out the potential tension between privacy and (some) nudges. They would surely agree that many acts of government, going beyond nudges, create such tension; both criminal and civil trials, not to mention the tax system, can raise serious privacy concerns. To come to terms with the underlying questions, it is necessary to specify both the nature and the weight of the privacy problem. We might doubt whether a home energy report, informing people about their levels of energy use, poses a truly grave risk to privacy, even if there is some danger that strangers might see the report. (On the list of current risks to privacy, home energy reports do not seem to belong on the top.) But Kapsner and Sandfuchs are certainly right to put the privacy problem on the table, even in apparently innocuous settings, and to suggest the importance of safeguards.

Guilhem Lecouteux

The whole idea of time inconsistency assumes that across time, we are dealing with a single person. At Time 1, Oscar makes a choice that ends up hurting him at Time 2 (say, retirement); a nudge at Time 1 might help Oscar to avoid a serious net welfare loss. But what if at Time 2, Oscar isn't the same person?

Citing Proust, Guilhem Lecouteux thinks that libertarian paternalism rests "on an implausible model of identity," and that a more plausible model causes some serious problems. Lecouteux believes that if we are to nudge young-Oscar in the interest of retired-Oscar, we must assume that all relevant Oscars have the "same true preferences"; that retired-Oscar's "revealed preferences correspond to young-Oscar's true preferences"; and that each Oscar is "rationally required to make time-consistent choices." Lecouteux thinks these assumptions are questionable. Retired-Oscar may be relevantly different from young-Oscar; he might regard as "unimportant" an effort that young-Oscar regarded as "intolerable." Retired-Oscar might not have the same true preferences as young Oscar. In any case, Oscar might not be "rationally required to discount his future utilities with a constant discount rate." At least this is so if we adopt a complex view of identity, associated with Derek Parfit, which does not see young-Oscar and retired-Oscar as "the same person, because they do not necessarily share the same memories, values or preferences."

If we adopt a Parfitian account based on psychological connectedness, rather than on a simple view of identity, we will see that "it is indeed not necessarily irrational to care less about one's further future if we know that one's identity is likely to evolve over time." Lecouteux concludes that if paternalism is to be justified, it is not to combat time inconsistency, but instead to promote autonomy. If people are shaped by factors of which they are not aware (framing effects and present bias might be such factors), then we might have a ground for intervention, perhaps by educating people about their biases, so as to increase the likelihood that they will be able to govern themselves.

Lecouteux deserves a great deal of credit for questioning behavioral claims about time inconsistency by reference to philosophical work on the nature of personal identity. His discussion is both dense and illuminating. Needless to say, he raises deep questions, which I cannot adequately engage here. But consider two tentative responses. First, we might accept some version of Lecouteux's view, but on welfarist grounds, we might think that if young-Oscar makes decisions that create significant net costs for Oscar, taken as an aggregation of Oscars over time, there is a real problem. On this view, of course, young-Oscar is not so different from (let's say) Otis, imposing costs on Owen, Odetta, Oliver, Omar, Orlando, Olive, Ophelia, and Otto. If Otis does that, then on standard grounds, it makes some sense to protect all those other O's, and perhaps nudges will do the trick. Lecouteux recognizes this point, and of course he thinks (with Parfit) that the relationship between young-Oscar and his successors is closer than that -- and if so, the welfarist argument for some nudging of young-Oscar, for the benefit of those successors, is strengthened rather than weakened.

The second response is to question whether and in what sense Parfit is right, and to wonder whether Lecouteux really wants to follow him. What does it even mean to say

that young-Oscar is a different person from retired-Oscar? Is this a claim of fact? If so, what kind of factual claim is it? Is it made out by virtue of (say) the fact that retired-Oscar has different memories and preferences? Why is that fact enough to make retired-Oscar a different person? Consider a competing view: In virtue of the *relevant* physical facts (for example, the same body, most importantly including the same brain), Oscar remains the same person over time. Or consider another view: The question whether Oscar is the same person over time is not, in the end, a question of fact. It is an interpretive question. For many purposes, including those raised by time inconsistency, whether Oscar remains the same person over time cannot be answered by reference to physical facts alone. The best way for Oscar, or for any human being, to make sense of his life is to understand himself as having a unitary identity. That form of sense-making is not inconsistent with any facts marshaled or claims made by Parfit or Lecouteux, and if I read him correctly, Lecouteux might well agree with it, emphasizing as he does that what matters is "the view of myself as an agent, as one who chooses and lives a particular life."

I tend to think that something like the third view is correct, and that it is ultimately consistent with behavioral concerns about time inconsistency. But we need not venture into the most contested theoretical waters to conclude that if young-Oscar makes decisions that seriously jeopardize the well-being of retired-Oscar, a nudge might be a good idea – even as we acknowledge that young-Oscar counts too, and that what matters is not just retired-Oscar, but Oscar's well-being over his lifetime.

Concluding Words

Nudges and choice architecture are hardly new; they are built into the fabric of human society, even when they seem invisible. Default rules – found in families, in customs, in norms, in law – cannot be blinked away, and they can greatly influence ultimate outcomes. What is new is sustained attention to a series of policy tools that often have large effects while preserving both agency and freedom of choice.

We are learning more, every day, about when, why, and how much nudges will matter. We are also learning a great deal about conceptual questions, about ethical issues, about likely public reactions to nudging, and about differences across cultures and nations. Future progress will depend on a high level of concreteness. We do not have to suffer from optimistic bias to believe that such progress is inevitable.

Reference List

Allcott, H. & Sunstein, C.R. (2015). Regulating Internalities. *Journal of Policy Analysis and Management* (forthcoming).

Barnhill, A. (2014). What is manipulation? In M. Weber & C. Coons (Eds.), *Manipulation: theory and practice* (pp. 51–72). Oxford, UK: Oxford University Press.

Bhargava, S., Loewenstein, G., Sydnor, J. (2015). Do employees make sensible health plan decisions? Evidence from a Menu with Dominated Options, forthcoming.

Bubb, R. & Pildes, R. H. (2014). How behavioral economics trims its sails and why, *Harvard Law Review*, *127*(6): 1593–1678.

Cadena, X. & Schoar, A. (2011). Remembering to pay? Reminders vs. financial incentives for loan payments, *National Bureau of Economics*, Working Paper No. 17020.

Camerer, C. F., Issacharof, S., Loewenstein, G., O'Donoghue, T., Rabin, M. (2003). Regulation for conservatives: behavioral economics and the case for asymmetric paternalism, *University of Pennsylvania Law Review*, *151*(3): 1211–1254.

Felsen, G., N. Castelo, and P. Reiner. 2013. Decisional Enhancement and Autonomy: Public Attitudes Toward Overt and Covert Nudges. *Judgment and Decision Making* 8(3)-202-213.

Goldin, J. (2015), Which way to nudge? Uncovering preferences in the behavioral age, doi: 10.2139/ssrn.2570930.

Jolls, C., Thaler, R. H., Sunstein, C. R. (1998). A behavioral approach to law and economics*, Stanford Law Review*, *50*(5): 1471–1550.

Halpern, D. (2015). *The nudge unit*. London: W.H. Allen (forthcoming).

Sharot, T. (2011). *The optimism bias: a tour of the irrationally positive brain*. New York: Knopf Doubleday.

Sunstein, C. R. (2015a). Nudging and choice architecture: ethical considerations. *Yale Journal on Regulation*, forthcoming.

Sunstein, C.R. (2015b). *Choosing not to choose*. Oxford: Oxford University Press.

Sunstein, C.R. (2015c). Fifty shades of manipulation. *Journal of Behavioral Marketing* (forthcoming).

Sunstein, C. R. (2014). *Why nudge?: The politics of libertarian paternalism*. New Haven: Yale University Press.

Sunstein, C. R. & Reisch, L. A. (2014). Automatically green: behavioral economics and environmental protection. *Harvard Environmental Law Review*, *38*(1): 127–158.

Sunstein, C. R. (2013). *Simpler: The future of government*. New York: Simon & Schuster.

Thaler, R. H. & Bernartzi, S. (2013). Behavioral economics and the retirement savings crisis, *Science*, *339*(6124): 1152–1153.

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth and happiness*. New Haven: Yale University Press.

Thaler, R. H., Sunstein, C. R., Balz, J. P. (2010), Choice architecture, doi: 10.2139/ssrn.1583509.

Ullman-Margalit, E. (1977). *The Emergence of Norms*. Oxford: Oxford University Press.

Willis, L. (2013). When nudges fail. *University of Chicago Law Review* 80: 1157-1227.

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, *61*(2): 341–355.

York, B. N. & Loeb, S. (2014). One step at a time: the effects of an early literacy text messaging program for parents of preschoolers, *National Bureau of Economics*, Working Paper No. 20659.

Author's Note: I am grateful to Adrien Barton, Tyler Cowen, Elizabeth Emens, Till Grune-Yanoff, Daniel Kahneman, Lucia Reisch, and Richard Thaler for extremely valuable comments and discussions.